Capture, representation, and "composition" of the instrumental gesture

Claude Cadoz, Christophe Ramstein

CAPTURE, REPRESENTATION, AND "COMPOSITION" OF THE INSTRUMENTAL GESTURE

Claude CADOZ, Christophe RAMSTEIN
A.C.R.O.E. / L.I.F.I.A.
Institut IMAG (Informatique et de Mathématiques Appliquées de Grenoble)
46 Av. Felix Viallet - 38000 GRENOBLE - FRANCE
Tel : 33 76 57 46 61 - Fax : 33 76 57 46 02

ABSTRACT : In designing and constructing a computer tool for musical creation, the INSTRUMENTAL GESTURE is especially relevant for real time control of the sound synthesis processes by simulation of insrtrumental mechanisms. This study is based on the implementation of a complete system to enable the capture and memorisation of the instrumental gesture during play, i.e. when the instrumentalist / instrument relationship is established. Memorised gestual actions enable two new problems to be tackled. These relate to the representation and the processing of instrumental gesture. The concepts that emerge lead to the elaboration of a "gesture editor".

The means available to man to make his presence felt to his outside environment are vocal and gestual. Verbal utterance is essentially used amongst his fellows whilst gesture is exclusively employed on objects, in space, and with or without tools. We can therefore say that in natural communication situations gesture is multiform. It is vigourous and motor, silent yet visible, minute but audible, and in all cases omnipresent.

Since the computer is playing an increasingly meaningful role in our environment, the obvious question that arises concerns the *communication* conditions between the former and the latter. Besides, the computer is both human and physical because it serves to manipulate the physical, as well as the abstract, world or yet again to communicate with other human beings through time and space.
Amongst the others, the question of "gesture and computer" thus exists in its own right.

In the field of computer music, this question becomes fundamentally and obviously significant, for in fact if one thing is vital to music it is the *instrumental gesture*. Moreover, if there is one area where gesture is multiform, varied, and rich it is definitely music.

The computer, and more generally electronic and electrotechnical technologies, have, historically, initally provided an objective grasp of the sound phenomenon, by authorising its capture, its tele-transmission, its processing, its memorisation ... and then came the image's turn.
Today's media are essentially "audio-visual".
In other less "mediatic" disciplines techniques relative to gesture processing have been developed for a long time now. Take, for example, *telemanipulation*. Thanks to appropriate organs, the gestual phenomenon in its turn can be captured, tele-transmitted, restituted, processed, and memorised.
It is quite clear that this is exactly what occurs when a gesture stroke is captured by an alphanumerical keyboard. In fact, this happened in a very reduced form with the Morse key manipulator, but what precisely had not then been understood, (and which still remains to be seen even today), is that gesture is a very subtle and rich phenomenon that requires amongst other things much more frequency bandpass than that offered by the current MIDI standard.

The instrumental gesture can become an object of research just like sound and image if we have suitable organs that are not only appropriate but also efficient enough.

We shall refer to these organs here as Gestual Transducers. They must be endowed with certain properties. In our opinion this makes for a research axis in its own right. As a result we have developed - (cf. The Modular Feedback Keyboard, Cadoz, Lisowski, Florens - in this conference) -

several devices given over to gestual communication. In particular, and thanks to force-feedback, these integrate tactilo-kinesthetic synthesis.
To begin with, these devices enable the instrumental gesture to *exist*. The geometry, articulatory properties, degrees of liberty, ... as well as the force-feedback have all been very carefully studied so that with the gesture can be executed in all its finesse and richness. The force-feedback plays a central role, for there is no fine gestual action that does develop the subtle senses of touch and kinesthetics.

We are going to introduce this research here. Our focus will be attached to the instrumental gesture as one of the objective elements, alongside the sound and the instrument (or their computer substitutes or representations), by which the musical composition process can operate. We shall then introduce the notions of **edition** and **editor** for the **instrumental gesture.**

A few guidelines for a gesture editor

The equipment that we use is the experimental device developed at the ACROE. It comprises :

- All the machines used for sound synthesis by physical models with the assistance of the CORDIS-ANIMA modular system.
- The sixteen-key version of the Modular Feedback Keyboard. This provides 16 signals corresponding to the key displacements (of around 3cm). These signals are sampled at a frequency that ranges from 300 to 1000 Hz, and coded over 12 bits.

The research we are presenting you here is based on the classical keyboard playing activity. In terms of the general instrumental gesture typology, this may be characterised as follows :

- *articulated digital gesture*, i.e. the combination of two types of action : the finger striking action, and the actual articulation of the fingers.
- The existence of an *equilibrium position for the keys* (that is identical for all the keys).

We cannot, in fact, establish an instrumental gesture typology without bearing in mind the organs it is based on.
The above case, which is the simplest and the most classical of all, was initially chosen as a point of departure from which we could set out the general theoretical problems as regards *coding*, *representing*, and *processing* the gestual phenomena. This basis will thus lead to a subsequent generalisation of sustained, or 'held-down' gestures to be applied to other morphologies.

Gesture Coding

We should distinguish between the need to code for economical transmission and memorisation (including archiving) of the gestual phenomena, and the requirement for an internal representation that is both adapted to the external representation (for the operator) and to the various base processing activities of the gestual phenomenon. Nonetheless, it would be preferable (but not necessarily feasible) if the coding in question were as near as possible to the internal representation.
In the case we are dealing with here, the actual question of the coding itself has been provisionally eluded : we use the 16 signals in their raw numeric form. It follows, therefore, that a transcoding phase must exist in the system.

Internal Representation

Prior to any "composition", there must be "decomposition", for the principle here is drawn from an existing material, i.e. the *gestual sequence* executed on the keyboard. This leads to the idea of an elementary gestual "event" which results from the hypothesis that any gestual sequence played on such a keyboard (with equilibrium position) can be decomposed into striking gestures that may be isolated on the various keys.

At this point, it should be pointed out that there is no identity between gesture and signal. In fact, the signal corresponds to the keys' movement which can be longer than the finger contact phase. Similarly, an entire part of the gesture (apart from the key contact) exists that is not sensed.

Hence, it is exaggerated to talk in terms of a gestual signal. Nevertheless, in the most general case, if we had a model of the operator's gestual organs as well as the mechanics of the instrument, we could infer from this signal more characteristic information of the gesture in itself (cf. research Gibet and Florens, ACROE - 87).

In the simple situation studied here, we focussed on the elementary event that represents the evolution of the key between the instant it is struck, and the moment, once released, when it returns to its rest position. We call this event a "SEGMENT".

The internal representation is based on this segment notion. Transcoding (segmentation) then consists in extracting the segments on each of the keys (lanes), by detecting the "pause" phases (the key in the rest position). In this way, we obtain a set of segments containing the key evolution data during its activity (sampling frequency and base quantification), to which we associate their action time and key number.

To cover more complex situations where the movements of the manipulated organs have several degrees of liberty (joy-stick ...), as well as those when several different devices (poly-instruments) are used together, we structured this information into "units" (relative to a set of identical organs), "channels" (relative to each organ), and "lanes" (relative to each degree of liberty for a given organ). A channel can be multidimensional (several lanes) and in this instance the preceding segment notion no longer applies to the lane, but to the channel : a signal can consist of a multidimensional signal.

Another structuration level is also introduced, corresponding this time to the articulation notion we have introduced above. We consider the groups of segments as associated, so as to constitute a "formula". Several imbrication levels are possible, so that we can make up the "formulae of formulae" in an arborescent representation. In this representation, a complete gestual sequence GI is described as a list of placed formulae (in the spatio-temporal field) :

GI ::= (Fp1 ... Fpn)
Fp ::= origine . attributs . Formule
Formule ::= (Fp1 ... Fpn) / Segment

The "attibutes" correspond to the different information relative to the formula. Their deifinition is an open definition.

External Representation

By this we mean the representation for the user. This representation is central to the tool, for in fact this must enable the operator to control the various manipulations and significative processing. We have adopted the following principles :

- Whatever the processing level, it must authorise access to the lowest representation level (in terms of signals in cartesian spatio-temporal coordinates ...).
- For each primitive event category we must look for a characterisation in "attributes". The latter are the manipulable traits in the signal that have a direct incidence on the function of the event. For example, for percussion gestures, these will be maximum intensity, the sharpness of the attack, etc. These attributes are not always the same, for this depends on the nature of the gesture. It should be noted that these are *gestual attributes* which are independent from the sound phenomenon that they provoke since the instrument they may be applied to can vary.

Once we have obtained a characterisation of the topology of the attributes, the visual representation then consists in associating graphic shapes with the same topology to these attributes. For example, for a percussive event, we can associate a line whose length is proportional to the intensity of the percussion, for a more categorial attribute (like the belonging to a given formula) we may associate a colour.

- Consider a representation focalised on the elementary event (the segment) that corresponds to a high level of detail for the segment, and a representation focalised on the articulation. The latter must show the relations between the events rather than the events themselves, but without

hiding any information concerning the events. Segment representation in the second is simplified, and is always obtained via the attributes.
The two representations are distinct, and can be called separately by the operator.

- No representation procedure is definitive. On the contrary, the system proposes a certain number of base representations, but remains available to any new definitions the operator may choose, or even define himself.

- Finally, the general representation mode of the gestual sequence uses "realistic" imagery. More precisely, we have defined a paradigm, (that we have coined the, "hiker paradigm") in the following way : the events (eventually mutiple) that arise in time are considered as objects set out in a space (a highway) and they take place in front of the operator (the hiker). The events (objects) are represented in cavalier perspective along this highway, and their spatial bulk is inversely proportionate to their temporal remoteness (or proportional to their "urgence"). This paradigm is very pertinent and operational since it effectively corresponds to the natural situation where objects are all that much smaller as they are far away in space and hence in time.
However, it has turned out that this realistic and concrete representation is better replaced by a more abstract form in 2D space, with time represented horizontally, beyond a certain degree of sequence complexity.
To accommodate for this we have implanted a system that allows the operator to switch from one to the other indifferently.

Processing

Processing, on the elementary level is shared into two categories : sequence processing and "articulation" processing.
As regards the second, this essentially, for the primitives, is a question of the well-known cut/paste/copy operations. They concern various entities ; segments, formulae. Classical mechanisms of designation, selection, cumulative selection, etc. have been set up. These primitives will be the basis for composition models that the operator himself can develop. For the moment, we have not implanted a system of "compositional" modelisation.
For the first category, elaborated processing models have been studied. However, we should first point out this corresponds to our approach on representation. On the one hand, the processing level for elementary information close to the "representation signal" remains accessible at all times, and on the other hand the definition of new models by the operator stays a possible option.

The specific models that we have developed and experimented, over and above the cartesian trace done by hand (with a mouse), introduce simplified operator and instrument mechanical models. This enables us to carry out transformations of the key displacement signal which retains the coherence of the latter. In this case, we have processing that authentically refers to gestual logic, and this will give rise to numerous developments in the course of our future research.

References

CADOZ, FLORENS, LUCIANI - 1981 - "Synthèse musicale par simulation des mécanismes instrumentaux. Transducteurs Gestuels Rétroactifs pour l'étude du jeu instrumental". Revue d'Acoustique N°59, pp. 279-292.
CADOZ, FLORENS, LUCIANI - 1984 - "Responsive Input Devices and sound synthesis by simulation of instrumental mechanisms : the CORDIS system". Computer Music Journal, 8 N°3, pp. 60-73.
CADOZ, LISOWSKI, FLORENS - 1990 - "Modular Feedback Keyboard". Computer Music Journal, 14N°2, pp. 47-51.
CADOZ, LISOWSKI, FLORENS - 1990 - "Modular Feedback Keyboard". ICMC Glasgow 1990.
CADOZ - 1988 - "Instrumental Gesture and Musical Composition". ICMC KÖLN 1988.
GIBET - 1987 - "Codage, Représentation, et Traitement du Geste Instrumental".
Thèse de doctorat de l'Institut National Polytechnique de Grenoble. Décembre 1987.

Aknowledgments : The research presented here is backed by the French MINISTRY for CULTURE.